Project 2

Evan Cheng

Abstract

The study by Hitsman et al. (2023) aimed to evaluate two smoking cessation methods—Behavioral Activation for Smoking Cessation (BASC) and Standard Behavioral Treatment (ST)—and the effectiveness of varenicline compared to a placebo for adults with Major Depressive Disorder (MDD). This randomized, placebo-controlled trial included approximately 300 participants and was conducted at two urban university research clinics in the United States.

The results demonstrated that varenicline significantly improved both short-term and long-term rates of quitting smoking compared to the placebo. Specifically, more participants were able to stop smoking and remain smoke-free over several months when using varenicline. However, BASC did not yield better results than ST in helping participants quit smoking. The research also identified issues with participants not fully engaging with or sticking to the prescribed behavioral treatments and medication plans. In simpler terms, many participants found it difficult to consistently follow the treatment activities and medication schedules prescribed in the study. Despite these challenges, the diverse backgrounds and mental health conditions of the participants highlighted the need for more tailored and intensive behavioral support along with medical treatment.

Building upon this foundational study, the present analysis seeks to explore how baseline characteristics may moderate the effects of behavioral treatment and pharmacotherapy on smoking cessation among adults with MDD. Using a predictive modeling approach with cross-validation, we employed logistic regression, Lasso, and Ridge regression models to assess interactions between treatment type and participant characteristics, examining both short- and long-term smoking cessation outcomes. Results from these models suggest that certain baseline variables, such as depression severity and nicotine dependence, interact significantly with treatment type, influencing cessation success rates. The findings underscore the importance of individualizing smoking cessation strategies for those with MDD and highlight the potential for using predictive modeling to optimize treatment approaches in clinical settings.

Introduction

The prevalence of tobacco use among individuals with major depressive disorder (MDD) presents a compelling challenge. While recent decades have seen a decline in smoking rates among the general population, those with MDD continue to smoke at significantly higher rates (Weinberger et al., 2020; Smith et al., 2019). This demographic not only tends to smoke more heavily but also experiences heightened nicotine dependence and more severe withdrawal symptoms, which can complicate cessation efforts (Hitsman et al., 2003).

Research has consistently demonstrated that smokers with depression benefit less from standard smoking cessation interventions compared to the general smoking population. This is partly due to psychological and neurobiological impairments such as diminished reward sensitivity and increased stress reactivity, which are common in depression and negatively impact smoking cessation outcomes (Hitsman et al., 2003; Cook et al., 2010).

Despite these challenges, specific treatments like Behavioral Activation (BA) and pharmacotherapy with varenicline have shown promise. BA is designed to increase engagement with rewarding activities and reduce avoidance behaviors, which are crucial barriers to cessation in depressed smokers (MacPherson et al., 2010). Varenicline, on the other hand, has been effective in reducing smoking reward, craving, and withdrawal symptoms, yet its efficacy may be influenced by concurrent behavioral therapies (Anthenelli et al., 2016; Evins et al., 2014).

The combination of pharmacotherapy and behavioral interventions tailored for this population has rarely been explored in a systematic manner. Most clinical trials have historically excluded individuals with mental health disorders, thereby leaving a significant gap in evidence-based strategies for this subgroup (Cinciripini et al., 2013; Haas et al., 2004). This study aims to address this gap by examining the efficacy of combined treatment approaches and identifying potential moderators and predictors of treatment outcomes in smokers with MDD. This will contribute to a more nuanced understanding of how different interventions can be optimized to support cessation in this high-risk population.

Data: Exploratory Analysis

The dataset employed in this project was derived from a study conducted by Hintsman et al. (2023), which investigated the impact of behavioral and pharmacological treatments on smoking cessation among individuals with major depressive disorder. The dataset includes a range of crucial variables for analyzing smoking abstinence outcomes. These variables encompass both treatment options, such as pharmacotherapy (with Varenicline) and psychotherapy, as well as a range of baseline characteristic variables. Characteristics include demographic information (age and sex, race/ethnicity), socio-economic status (income and education level), and detailed smoking behavior metrics (e.g., FTCD score and cigarettes per day). Additionally, psychological profiles such as the Beck Depression Inventory score, Pleasurable Events

Scales, and Anhedonia scores are included to assess the individuals' mental health context, which can pontetially influence cessation success. This dataset provides the opportunity to explore how baseline variables might moderate the success of cessation treatments. A detailed data summary table is provided below for reference (Table 1. Summary Table).

The current dataset includes a diverse range of participants in terms of age, sex, and race. Most participants are within the middle-age bracket. This age distribution is relevant due to the interaction between age, smoking habits, and responses to cessation (Figure 1. Age Distribution by Race). Regarding sex, the data includes 165 male participants and 135 female participants, resulting in a near fifty-fifty split.The racial demographics are primarily Black/African American and Non-Hispanic White. While Hispanic and other races are included, they represent only an insignificant portion of the dataset. This could be a limitation, as culture and genetics can play a crucial role in smoking habits and responses to cessation.

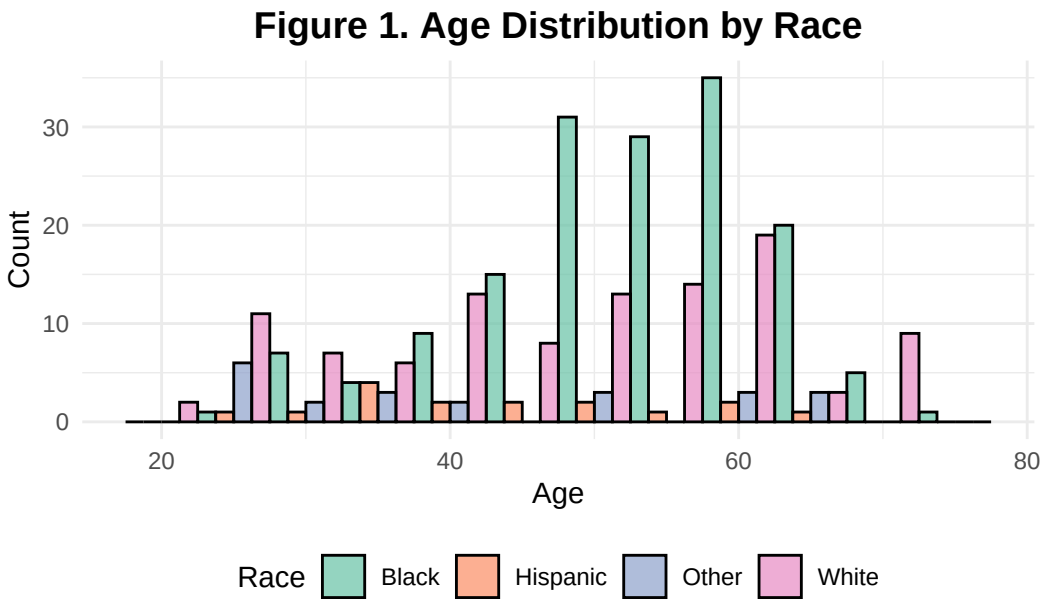
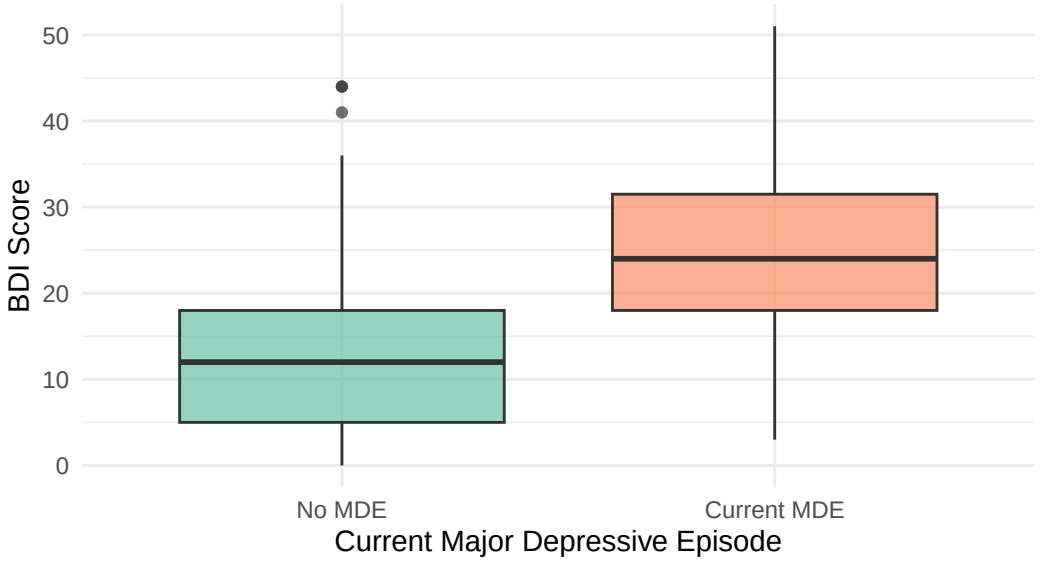


Figure 1. Age Distrubution By Race shows the varying density distributions of age

Since the current study focuses on the effect of Varenicline among individuals with MDD, depression measurement is a critical variable, as It should be randomly assigned to each treatment group to ensure appropriate variability. The box-plots shows that treatment assignment is systematically balanced across the Behavioral Activation (BA) and Standard Treatment (ST) groups (Figure 2. BDI Scores By Treatment Group). This balance is also shown in the Varenicline and placebo groups (Figure 3. BDI Scores By Medication vs. Placebo). In other words, each treatment group consist diverse baseline characteristics. Such stratification ensures that demographic differences do not confound treatment comparisons. Key variables,

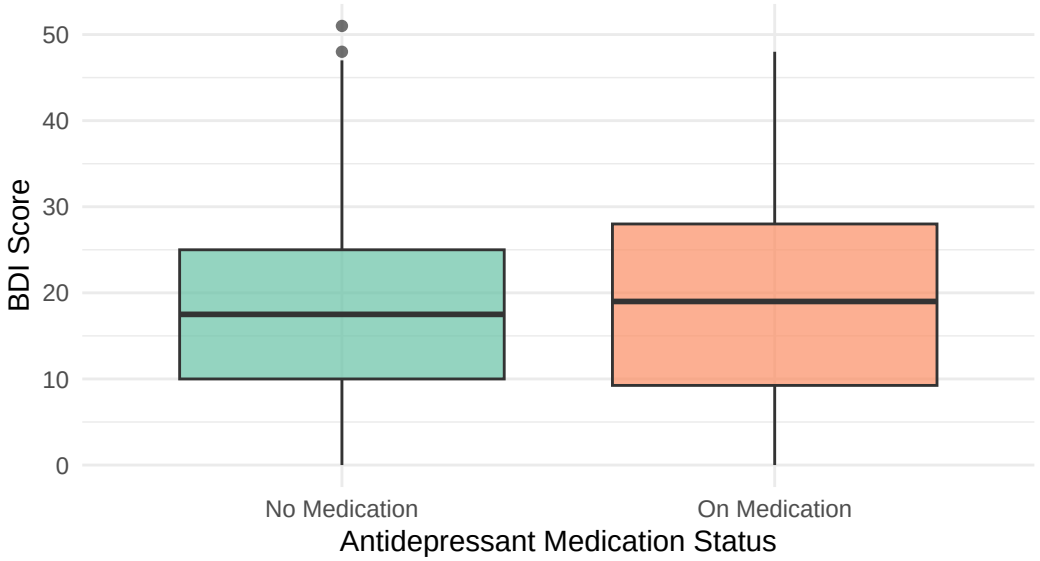
like age and Beck Depression Inventory scores, display stable standard deviations. This stability in variance is essential for the validity of inferential statistics used in evaluating treatment efficacy, ensuring that the model can lead to robust interpretations.

Figure 2. BDI Scores by Treatment Group



Distribution of Beck Depression Inventory (BDI) scores compared between particip

Figure 3. BDI Scores by Medication vs. Placebo



Comparison of Beck Depression Inventory (BDI) scores between participants takir

Table 1. Missing Data Summary

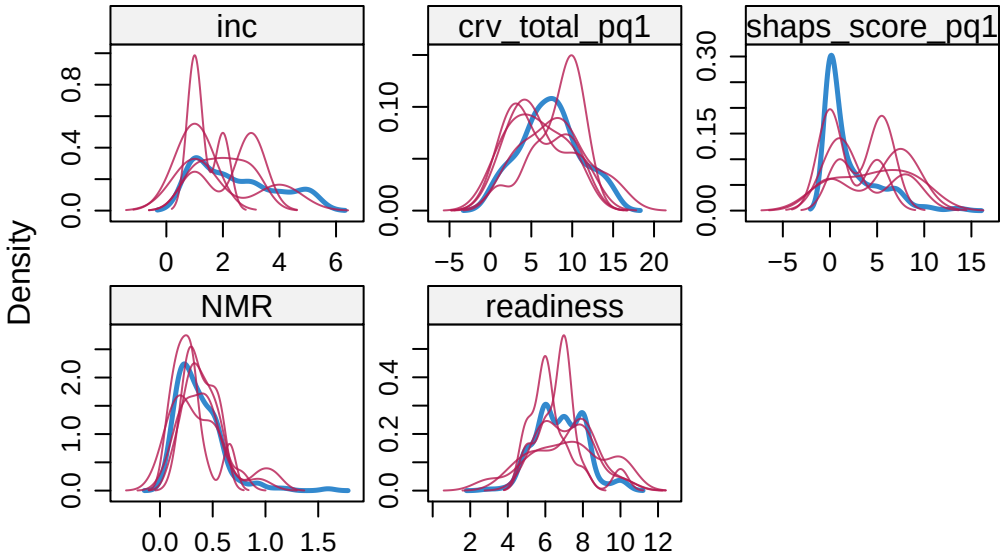
Variables	Missing Values	Percentage Missing
Income	3	1.00
FTCD Score at Baseline	1	0.33
Cigarette Reward Value at Baseline	18	6.00
Anhedonia	3	1.00
Nicotine Metabolism Ratio	21	7.00
Exclusive Mentholated Cigarette User	2	0.67
Baseline Readiness to Quit Smoking	17	5.67
Total	65	21.67

Missing Data

The dataset for this study, as shown by the missing data visualization, displays a relatively moderate level of missing data across several key variables (Figure 4. Missing Data Summary). However, notable exceptions include Nicotine Metabolism Ratio and Cigarette Reward Value Perception which exhibit the highest occurrences of missing entries. The presence of missing data require careful consideration of our approach.

Given the substantial proportion of missing data (21.667% overall) across several key variables, with particularly notable missingness in Nicotine Metabolism Ratio (7%) and Cigarette Reward Value at Baseline (6%), implementing multiple imputation is the most appropriate approach for this analysis. This decision is supported by two main factors: first, the cumulative impact of missing data across variables could significantly reduce the effective sample size if complete case analysis were used; second, the moderate size of our dataset means we cannot afford to lose statistical power by removing cases with missing values.

Multiple imputation provides a robust solution that allows us to retain all observations while accounting for uncertainty in the imputed values. This is especially important for key variables like Nicotine Metabolism Ratio and Cigarette Reward Value at Baseline, which are crucial for understanding smoking cessation patterns. To maintain the integrity of our cross-validation process and prevent any potential data leakage, the imputation will be performed separately on the training and testing datasets. This separation ensures that information from the test set does not influence the imputation of the training set and vice versa, thereby providing a more reliable assessment of our model's performance.



Based on the density plots from our multiple imputation analysis, the imputed values (red) generally maintain similar distributions to the observed data (blue) across key variables. The imputations for NMR and readiness to quit smoking show particularly good alignment with the observed distributions, suggesting reliable imputation quality. The cigarette reward value and SHAPS score imputations demonstrate some variability across imputed datasets but maintain plausible distributions relative to the observed data. Incomeshows more variation in the imputed values, which is expected given its categorical nature and the complexity of socioeconomic factors. These diagnostic plots suggest that our multiple imputation approach has produced reasonable estimates while preserving the key features and relationships in the original data.

Method

The goal of the current study is to analyze potential moderators and predictors influencing smoking cessation outcomes, distinguishing between interaction effects in behavioral treatments and direct baseline influences. Logistic regression will be employed for both moderation and predictive models, since it's commonly used for analyzing binary outcomes. Lasso and Ridge regression will also be implemented to manage multicollinearity and reduce model complexity. Lasso is particularly valued for its ability to perform variable selection by shrinking less significant coefficients to zero. This way, the model can focus on the most impactful interactions. Conversely, Ridge regression will maintain the contributions of all variables without exclusion; mainly, to compare performance with a Lasso model.

Cross-validation will be integrated during model development to ensure robustness and generalizability. A k-fold cross-validation allow for validation by splitting the data into multiple subsets, which assess each model's performance across different segments. This approach can further reduce overfitting, and offers a accurate estimate of model performance. Each subset will be used as a validation set once, while the remaining subsets serve as training data. This ensures every data point contributes to model validation. Using cross-validation not only enhances the credibility of model findings but also strengthens the model against variations in data.

Models then will be evaluated using the Area Under the Receiver Operating Characteristic (AUC-ROC) curve. It is used to assess the model's discriminatory ability across various thresholds. This evaluation will be complemented by the examination of confusion matrices since it provides insights into the models' performance by highlighting the occurrence of false positives and false negatives. These metrics will be instrumental in comparing the effectiveness of interaction-focused moderation models against the direct influence of baseline predictors in the predictive models, thereby illuminating the differential impacts of treatment variables and baseline characteristics on cessation outcomes.

Results

In the Lasso regression model, a significant number of coefficients were reduced to zero, highlighting Lasso's strength in simplifying the model and selecting only the most relevant features. This capability is especially valuable in the current settings, where identifying impactful variables is critical for targeted interventions. For instance, "Age" and "FTCD Score at Baseline" showed negative coefficients in both models, meaning that older individuals and those with higher nicotine dependence are less likely to quit smoking. In other words, the findings suggest that older age and a higher dependence on nicotine may serve as barriers to cessation, possibly due to the difficulty older adults face in breaking established habits or the stronger physiological hold nicotine has on those with higher dependence.

Variable	LASSO	Ridge	Logistic
(Intercept)	-1.8917871	-0.8797308	-3.4666446
inc	-0.0944827	-0.2310762	-0.3478421
ftcd.5.mins	-0.2232236	-0.7018093	-1.5821013
bdi_score_w00	0.0978864	0.0792587	0.1385432
otherdiag	0.6589643	0.9222647	1.3924998
id	0.0000000	-0.0020997	-0.0030720
abst	0.0000000	-0.1714032	-0.0466299
Var	0.0000000	0.0264815	0.0563471
BA	0.0000000	-0.1782070	-0.2343528
age_ps	0.0000000	-0.0106521	-0.0185743

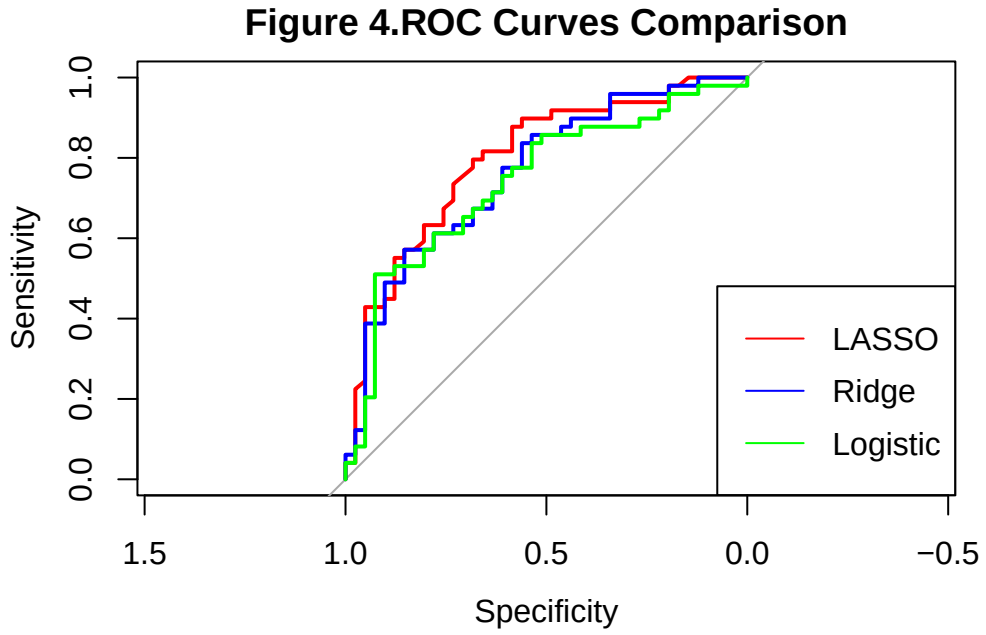
Variable	LASSO	Ridge	Logistic
sex_ps	0.0000000	0.4178731	0.8061799
NHW	0.0000000	0.0101769	1.4075599
Black	0.0000000	0.1824905	1.7566159
Hisp	0.0000000	-0.3596502	-13.6319575
edu	0.0000000	0.2277895	0.5269654
ftcd_score	0.0000000	0.1285657	0.3615827
cpd_ps	0.0000000	-0.0105080	-0.0224314
crv_total_pq1	0.0000000	-0.0042696	-0.0224962
hedonsum_n_pq1	0.0000000	-0.0134755	-0.0284367
hedonsum_y_pq1	0.0000000	-0.0082072	-0.0192144
shaps_score_pq1	0.0000000	-0.0388342	-0.1742844
antidepmed	0.0000000	-0.1960684	-0.3861659
NMR	0.0000000	-0.5759445	-1.3567727
Only.Menthol	0.0000000	-0.2347380	-0.4856331
readiness	0.0000000	-0.0581979	-0.0935086
raceHispanic	0.0000000	-0.0832489	14.1951963
raceOther	0.0000000	-0.5845726	0.0000000
raceWhite	0.0000000	0.0106973	0.0000000

On the other end, Ridge regression retains all variables in the model but shrinks their coefficient values, allowing for a more nuanced understanding of each predictor's impact without excluding any. This was observed with the BDI Score, where Ridge showed a moderated influence of baseline depression on quitting outcomes. In contrast, Lasso's treatment of this variable fluctuated between positive and negative impacts. In other words, Ridge suggests a more balanced effect of depression on quitting success, while Lasso implies that depression could have varied effects, perhaps affecting patient motivation and response to treatment in complex ways.

Both models achieved an Area Under the Curve (AUC) of 0.73, demonstrating strong predictive accuracy and a solid capacity to differentiate between individuals likely and unlikely to quit smoking by the treatment's end. In other words, both models performed effectively in forecasting cessation outcomes, despite their differences in how they handle predictor variables. This equivalence in AUC, even with variations in coefficient magnitudes and significance, suggests that both Lasso and Ridge offer robust and reliable frameworks for understanding and predicting smoking cessation outcomes.

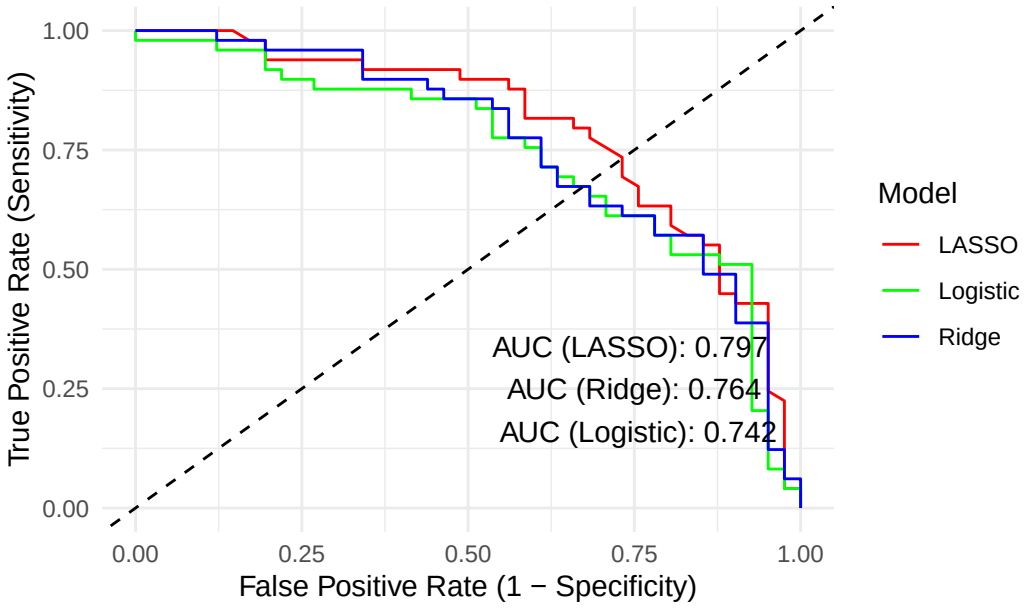
The logistic regression model, on the other end, offered valuable insights into predictors linked to smoking abstinence. Indicated by the coefficients and significance values, baseline variables such as FTCD score ($p = .020$) and time to first cigarette ($p = .023$) shown as statistically significant. This suggest that baseline nicotine dependence levels significantly influence abstinence outcomes. In other words, higher dependence levels appear to reduce the likelihood of

successfully quitting.



The model's Area Under the Curve (AUC) score was 0.519, which is lower than both Lasso and Ridge, indicates limited ability to distinguish between abstinent and non-abstinent individuals. This low AUC suggests that logistic regression, although helpful in interpreting the individual effects of predictors, may lack the the ability to capture the complex moderator effect. In simple terms, the logistic regressionmay not model the relationships as effectively as methods like Lasso or Ridge, which explains why some variables in this model show weaker predictive power.

Figure 5. ROC Curves Comparison



In summary, the comparative analysis of the Lasso, Ridge, and Logistic Regression models underscores the critical role of selecting suitable modeling techniques that align with the research goals and the characteristics of the data. Each model offers unique insights into the factors affecting smoking cessation, with Lasso and Ridge models pinpointing key predictors through regularization, while the Logistic Regression model provides a more traditional, interpretable framework. Together, these models reveal the complex, multifaceted nature of the cessation process. In clinical settings, insights gained from these models could guide the design of more personalized and effective interventions by focusing on the most impactful predictors identified across models, ultimately improving the success of smoking cessation programs.

Given the results above, the Lasso model provided a simplified yet accurate prediction framework that effectively highlights significant predictors in smoking cessation for individuals with major depressive disorder (MDD). Among these, the nicotine metabolism ratio stood out as a critical factor, suggesting individuals with higher metabolism rates may need custimized treatment to increase cessation success. Additionally, the model identified the difference between current and past MDD as a crucial variable, indicating that active symptoms may have a distinct impact on the ability to quit smoking compared to previous symptoms. These findings provided by the Lasso model can inform more personalized and effective treatment protocols in the clinical settings.

Limitation

Limitation of the Lasso model in this context can be its tendency to zero out coefficients for variables viewed as less important. This can lead to an oversimplified model. While this regularization technique is beneficial for reducing complexity and multicollinearity, it might exclude potentially relevant predictors that have subtle but meaningful impacts on smoking cessation, such as cigarette per day or BDI score. Given the complexity of behavioral health, factors that are slightly less predictive but clinically significant could be overlooked.

The Ridge model shrinks all coefficients toward zero rather than eliminating them entirely. While this helps retain all predictors and mitigates the risk of missing subtle effects, it also tends to dilute the impact of significant predictors, especially in cases where there is high multicollinearity among variables. This can lead to a model that captures general patterns but might lack the specificity needed to identify the most influential factors in smoking cessation. In this study's context, where certain psychological or demographic factors may present distinct impacts, the Ridge model can underrepresent those effects, which provides a more generalized rather than targeted understanding.

Logistic Regression is widely used and highly interpretable, but has limitations in handling complex relationships and interactions among variables. In this analysis, factors such as behavioral treatment, pharmacotherapy, and baseline demographics likely interact in non-linear ways that logistic regression struggle to capture. Additionally, logistic regression does not inherently account for multicollinearity. This could potentially bias the results, especially for predictors that are highly correlated, and reducing the robustness of model insights

Reference

Anthenelli, R. M., Benowitz, N. L., West, R., St Aubin, L., McRae, T., Lawrence, D., & Evins, A. E. (2016). Neuropsychiatric safety and efficacy of varenicline, bupropion, and nicotine patch in smokers with and without psychiatric disorders (EAGLES): A double-blind, randomised, placebo-controlled clinical trial. *The Lancet*, *387*(10037), 2507–2520. https://doi.org/10.1016/S0140-6736(16)30272-0

Cinciripini, P. M., Robinson, J. D., Karam-Hage, M., Minnix, J. A., Lam, C., & Versace, F. (2013). Effects of varenicline and bupropion sustained-release for smoking cessation in adults with a history of major depressive disorder (MDD): A randomized, double-blind, placebo-controlled study. *Journal of Clinical Psychiatry*, *74*(2), 119–126. https://doi.org/10.4088/JCP.12m07871

Cook, J. W., Spring, B., McChargue, D. E., & Hedeker, D. (2010). Effects of anhedonia on days to relapse among smokers with a history of depression: A brief report. *Nicotine & Tobacco Research*, *12*(9), 978–982. https://doi.org/10.1093/ntr/ntq127

Evins, A. E., Cather, C., & Culhane, M. A. (2014). Tobacco cessation treatment for smokers with mental illness or addiction. *Journal of Clinical Psychiatry*, *75*(8), 1039–1047. https://doi.org/10.4088/JCP.13m08783

Haas, A. L., Muñoz, R. F., Humfleet, G. L., Reus, V. I., Hall, S. M., & Munoz, R. F. (2004). Influences of mood, depression history, and treatment modality on outcomes in smoking cessation. *Journal of Consulting and*

Hitsman, B., Borrelli, B., McChargue, D. E., Spring, B., & Niaura, R. (2003). History of depression and smoking cessation outcome: A meta-analysis. *Journal of Consulting and Clinical Psychology*, *71*(4), 657–663. https://doi.org/10.1037/0022-006X.71.4.657

Lerman, C., Audrain, J., Orleans, C. T., Boyd, R., Gold, K., Main, D., Caporaso, N., & Shields, P. G. (2007). Investigation of mechanisms linking depressed mood to nicotine dependence. *Addictive Behaviors*, *32*(1), 278–287. https://doi.org/10.1016/j.addbeh.2006.03.034

MacPherson, L., Tull, M. T., Matusiewicz, A. K., & Lejuez, C. W. (2010). Family smoking environment and cigarette refusal skills in adolescents: The roles of parents, friends, and siblings. *Addictive Behaviors*, *35*(9), 800–805. https://doi.org/10.1016/j.addbeh.2010.04.003

Smith, P. H., Mazure, C. M., & McKee, S. A. (2019). Smoking and mental illness in the US population. *Tobacco Control*, *28*(6), 648–653. https://doi.org/10.1136/tobaccocontrol-2018-054265

Weinberger, A. H., Gbedemah, M., Martinez, A. M., Nash, D., Galea, S., & Goodwin, R. D. (2020). Trends in cigarette use among adults with major depressive episodes in the United States, 2005 to 2016. *JAMA Psychiatry*, *77*(9), 893–896. https://doi.org/10.1001/jamapsychiatry.2020.1345